

PATENT CLAIMS

1. A method for analyzing fundamental frequency information contained in voice samples, characterized in that it comprises at least:

- a step (2) for the analysis of the voice samples grouped together in frames in order to obtain, for each sample frame, spectrum-related information and information relating to the fundamental frequency;
- a step (20) for the determination of a model representing the common characteristics of the spectrum and fundamental frequency of all samples; and
- a step (30) for the determination of a fundamental frequency prediction function exclusively according to spectrum-related information on the basis of said model and voice samples.

2. The method as claimed in claim 1, characterized in that said analysis step (2) is adapted to supply said spectrum-related information in the form of cepstral coefficients.

3. The method as claimed in any one of claims 1 or 2, characterized in that said analysis step (2) comprises:

- a sub-step (4) for modeling voice samples according to a sum of a harmonic signal and a noise signal;
- a sub-step (5) for estimating frequency parameters, and at least the fundamental frequency of the voice samples;
- a sub-step (6) for synchronized analysis of the fundamental frequency of each sample frame; and
- a sub-step (7) for estimating the spectral parameters of each sample frame.

4. The method as claimed in any one of claims 1 to 3, characterized in that it furthermore comprises a step (10) for normalizing the fundamental frequency of each sample frame in relation to the mean of the fundamental frequencies of the analyzed samples.

5. The method as claimed in any one of claims 1 to 4, characterized in that said step (20) for the determination of a model corresponds to the determination of a model by mixing Gaussian densities.

5 6. The method as claimed in claim 5, characterized in that said model determination step (20) comprises:

- a sub-step (22) for determining a model corresponding to a mixture of Gaussian densities; and
 - a sub-step (24) for estimating the parameters of
- 10 the mixture of Gaussian densities on the basis of the estimation of the maximum resemblance between the spectral information and the fundamental frequency information of the samples and of the model.

7. The method as claimed in any one of claims 1 to 15 6, characterized in that said step (30) for the determination of a prediction function is implemented on the basis of an estimator of the implementation of the fundamental frequency, knowing the spectral information of the samples.

20 8. The method as claimed in claim 7, characterized in that said step (30) for determining the fundamental frequency prediction function comprises a sub-step (32) for determining the conditional expectation of the implementation of the fundamental frequency, knowing

25 the spectral information, on the basis of the a posteriori probability that the spectral information is obtained on the basis of the model, the conditional expectation forming said estimator.

9. A method for the conversion of a voice signal 30 pronounced by a source speaker into a converted voice signal whose characteristics resemble those of a target speaker, comprising at least:

- a step (50) for determining a function for the transformation of spectral characteristics of the
- 35 source speaker into spectral characteristics of the target speaker, implemented on the basis of voice samples of the source speaker and the target speaker; and

- a step (70) for transforming spectral information of the voice signal of the source speaker to be converted with the aid of said transformation function,

5 characterized in that it furthermore comprises:

- a step (60) for determining a fundamental frequency prediction function exclusively according to spectrum-related information for the target speaker, said prediction function being obtained with the aid of
10 an analysis method as claimed in any one of claims 1 to 8; and

- a step (80) for predicting the fundamental frequency of the voice signal to be converted by applying said fundamental frequency prediction function
15 to said transformed spectral information of the voice signal of the source speaker.

10. The method as claimed in claim 9, characterized in that said step (50) for determining a transformation function is implemented on the basis of
20 an estimator of the implementation of the target spectral characteristics, knowing the source spectral characteristics.

11. The method as claimed in claim 10, characterized in that said step (50) for determining a
25 transformation function comprises:

- a sub-step (52) for modeling the source and target voice samples according to a sum model of a harmonic signal and a noise signal;

- a sub-step (54) for aligning the source and
30 target samples; and

- a sub-step (56) for determining said transformation function on the basis of the calculation of the conditional expectation of the implementation of the target spectral characteristics, knowing the
35 implementation of the source spectral characterizations, the conditional expectation forming said estimator.

12. The method as claimed in any one of claims 9 to 11, characterized in that said transformation

function is a spectral envelope transformation function.

13. The method as claimed in any one of claims 9 to 12, characterized in that it furthermore comprises a
5 step (65) for analyzing the voice signal to be converted, adapted to supply said spectrum-related information and information relating to the fundamental frequency.

14. The method as claimed in any one of claims 9 to 13, characterized in that it furthermore comprises a
10 synthesis step (90), enabling the formation of a converted voice signal at least on the basis of the transformed spectral information and the predicted fundamental frequency information.

15 15. A system for converting a voice signal (110) pronounced by a source speaker into a converted voice signal (120) whose characteristics resemble those of a target speaker, said system comprising at least:

- means (104) for determining a function for
20 transforming spectral characteristics of the source speaker into spectral characteristics of the target speaker, receiving, at their input, voice signals of the source speaker (100) and of the target speaker (102); and

25 - means (114) for transforming spectral information of the voice signal (110) of the source speaker to be converted by applying said transformation function supplied by the means (104),

characterized in that it furthermore comprises:

30 - means (106) for determining a fundamental frequency prediction function exclusively according to spectrum-related information for the target speaker, adapted for the implementation of an analysis method as claimed in any one of claims 1 to 8, on the basis of
35 voice samples (102) of the target speaker; and

- means (116) for predicting the fundamental frequency of said voice signal to be converted (110) by applying said prediction function determined by said means (106) for determining a prediction function to

said transformed spectral information supplied by said transformation means (114).

16. The system as claimed in claim 15, characterized in that it furthermore comprises:

5 - means (112) for analyzing the voice signal to be converted (110), adapted to supply, at their output, spectrum-related information and information relating to the fundamental frequency of the voice signal to be converted; and

10 - synthesis means (118) enabling the formation of a converted voice signal on the basis of at least the transformed spectral information supplied by the means (114) and the predicted fundamental frequency information supplied by the means (116).

15 17. The system as claimed in either of claims 15 and 16, characterized in that said means (104) for determining a transformation function are adapted to supply a spectral envelope transformation function.

20 18. The system as claimed in any one of claims 15 to 17, characterized in that it is adapted for the implementation of a voice conversion method as claimed in any one of claims 9 to 12.